Algebraic Combinatorics

on

Words

M. Lothaire

May 30, 2001

Contents

Preface

Combinatorics on words is a field that has grown separately within several branches of mathematics, such as number theory, group theory or probabilities, and appears frequently in problems of theoretical computer science, as dealing with automata and formal languages.

A unified treatment of the theory appeared in Lothaire's 'Combinatorics of Words". Since then, the field has grown rapidly. This book presents new topics of combinatorics on words.

Several of them were not yet ripe for exposition, or even not yet explored twenty years ago. The spirit of the book is the same, namely an introductory exposition of a field, with full proofs and numerous examples, and further developments deferred to problems, or mentioned in Notes.

This book is independent of Lothaire's first book, in the sense that no knowledge of the first volume is assumed. In order to avoid repetitions, some results of the first book, when needed here, are explicitly quoted, and are only referred for the proof to the first volume.

This volume presents, compared to the previous one, two important new features. It is first of all a complement in the sense that it goes deeper in the same direction. For example, the theory of unavoidable patterns (Chapter 3) is a generalization of the theory of square-free words and morphisms. In the same way, the chapters on statistics on words and permutations (Chapters 10 and 11) are a continuation of the chapter on transformations on words of the previous volume. But this volume is also a complement in the sense that it presents aspects of Combinatorics on Words that had not been treated in the previous one. For example, the plactic monoid is presented here although it had not been mentioned at all in the previous volume. The same holds for several topics connected with symbolic dynamics, namely Sturmian words or beta-expansions.

Let us now describe more in detail the content of this volume. Most of the basic facts needed are given in Chapter 1 "Finite and Infinite Words", written by Jean Berstel and Dominique Perrin. This chapter also contains basic concepts on symbolic dynamical systems. Unavoidable sets of words are studied at the end of this chapter. They are considered again in Chapter 3.

Chapter 2 "Sturmian Words", written by Jean Berstel and Patrice Séébold, is a systematic exposition of a family of infinite words that have minimal complexity. These words share a number of extremal properties and can be defined in several quite different ways. After a treatment of these properties, morphisms preserving Sturmian words are characterized. A strong relationship to the continued fraction expansion of irrational numbers is established.

Chapter 3 "Unavoidable Patterns", is written by Julien Cassaigne. A pattern is unavoidable if there exist infinitely many words that do not encounter this pattern. This is the generalization of square-free words (covered in Lothaire's first book). The algorithm of Zimin for testing whether a pattern is unavoidable is given. When the alphabet is fixed, one gets a hierarchy of avoidability (squares are 3-avoidable but not 2-avoidable). Some results concerning this hierarchy are derived.

Chapter 4 "Sesquipowers" is written by Aldo De Luca and Stefano Varricchio. Sesquipowers can be defined by bi-ideal sequences. These sequences have interesting combinatorial properties, with links to recurrence and n-divisions. From these an improvement of an important theorem by Shirshov is obtained. Regularities of Coudrain and Schützenberger, and of Shirshov, can be presented in a unified way. Applications to finiteness conditions in semigroups are given.

Chapter 5 "The Plactic Monoid" is written by Alain Lascoux, Bernard Leclerc and Jean-Yves Thibon. The plactic monoid is an algebraic structure that takes into account most of the combinatorial properties of Young tableaus. The starting point of the theory is Schensted's algorithm. The defining relations of the plactic monoid were determined by D. Knuth. Applications include the Littlewood-Richardson rule, a combinatorial description of the Kostka-Foulkes polynomials, a noncommutative version of the Demazure character formula, and of the Schubert polynomials. Quite recently, combinatorics of Young tableaus were related to quantum groups.

Chapter 6 "Codes", written by Véronique Bruyère, is concerned with several kinds of codes, in relation to the so called defect theorem. The defect effect still holds if the set is not an ω-code. A remarkable phenomenon appears when, for a finite code X, neither X nor its reversal $\widetilde{X}$ is an ω-code. In this case, the n elements of X can be expressed as a product of $n-2$ words. The chapter ends with a short and elementary proof of a result of Schützenberger stating that a finite maximal code X that is also an ω-code is prefix.

Chapter 7 "Numeration Systems", written by Christiane Frougny, deals with the various ways to write integers, reals, and complex numbers in positional number systems. Finite automata may exist to perform arithmetic operations, such as addition, and also to compute some standard representation. A special class of representations, called β-expansions, has several interesting properties related to symbolic dynamical systems. Generalizing the notion of base leads to number systems with respect to a sequence of numbers, such as the Fibonacci numbers. Numeration systems for complex numbers, without sign, and without separating real and imaginary parts, are considered at the end of the chapter.

Chapter 8 "Periodicity", written by Filippo Mignosi and Antonio Restivo, considers periods of various kinds in finite and infinite words. Repetitions may be of rational (not only integer) order. The golden ration appears to be an extremal value for periodicity in words. An important topic is the relation between local and global periodicity. Criteria for infinite words to be periodic are given next. Again, the golden ratio plays a central role.

The aim of Chapter 9 "Centralizers of Noncommutative Series and Polynomials", written by Christophe Reutenauer, is to give a self contained proof of Cohn and Bergman's centralizer theorems. These are analogues, in polynomials and series, of the well-known fact that two commuting words are powers of a third word. The proofs use noncommutative Euclidean division, the result that the centralizer of a noncommutative polynomial is integrally closed in its field of fractions, its embeddability in a one variable polynomial ring, and a characterization of free subalgebras of a one variable polynomial algebra. In addition, a defect theorem is shown to hold for two noncommutative polynomials.

Chapter 10 "Transformations on Words and q-Calculus", written by Dominique Foata and Guo-Niu Han, deals with statistics on words. There are several relevant statistics, such as the number of descents, of excedances, the major index, and the Denert statistics. MacMahon had already calculated the distributions of the early statistics. All calculations are presented here in a unified way. The second part is devoted to the derivation of an algorithm, that involves the introduction of commutation rules on biwords, and serve to the construction of two bijections. The chapter concludes with the proof of equidistribution properties.

Chapter 11 "Statistics on Permutations and Words" is written by Jacques Désarménien. It starts with the so called shape of a word, computes a statistics on shapes, considers inversion of permutations with a given shape. Lyndon words are related to cycles of permutations.

Chapter 12 "Makanin's Algorithm", written by Volker Diekert, is a self-contained exposition of the famous theorem of Makanin stating that it is decidable whether a set of equations in words has a solution. The first step towards Makanin's result is to bound the exponent of periodicity. Next, the problem is transformed to systems of boundary equations. This leads to a geometric reflection of the problem. An upper bound for the exponent of periodicity yields an upper bound on the length of convex chains. This in turn leads to an upper bound on the number of boundary equations. Then transformation rules are defined which either lead to a solution, or introduce additional boundary equations. Since their number is bounded, this procedure eventually stops.

Chapter 13 "Independent Systems of Equations" written by Tero Harju, Juhani Karhumäki and Wojciech Plandowski, is concerned with the existence of a notion of dimension for a set of words. A good example is the defect theorem already considered earlier. Another result is the compactness property (also known as Ehrenfeucht's conjecture) stating that every independent set of equations in words is finite. Existence of independent system of equations, together with bounds on their size, are given. Although the problem generalizes in a natural way to all semigroups, the compactness property does not hold in all semigroups. Varieties of semigroups with that property are characterized in terms of ascending chains of congruences.

Each chapter of this book can be read independently of the others, in the sense that there is no logical dependency of the results of one chapter with those of another one. The introductory chapter (Chapter 1) is an exception, however. It contains definitions and results used in the rest of the volume and it has been designed as a reference for the other ones. Each of the chapters can be used for a separate graduate course or seminar. The necessary mathematical background does not exceed a general undergraduate level.

A word about the process which gave rise to this volume. The authors of the previous volume have accepted to serve as a steering committee. The set of authors is to a large extent different from the previous one. On several occasions, the whole content of the book has been presented in seminars. This includes a special session of the Lotharingian seminar held in Bellagio in October 1996.

Finally, a third volume of combinatorics on words is in preparation. It is focused on applications. This includes natural language processing, text algorithms, fractals and tilings and bioinformatics.

The authors acknowledge helpful discussions and comments with a great number of colleagues. Among them are Laurence Bartz, Paul Cohn, Clelia de Felice, Jeanne Devolder, Georges Hansel, Jacques Justin, Michel Koskas, Michel Latteux, Jean Mairesse, Yuri V. Matiyasevich, Anca Muscholl, Bruno Pettazoni, Gwénaël Richomme, Klaus U. Schulz, Stephanie van Willigenbourg.

Jean Berstel
Dominique Perrin

Marne-la-Vallée, May 30, 2001